IRNC NetSage – Use Cases and Scalability

Doug Southworth, Jennifer M. Schopf

November, 2020

What NetSage Does Best

- Answers questions asked by network engineers and network owners
- Human-readable summaries and patterns
- Gives people the higher level pattern so they can narrow down a time frame and then use local tools that have more detail
- Simplifies and makes accessible basic data

Monitoring using NetSage

- NetSage advanced measurement services for R&E data traffic
  - Better understanding of current traffic patterns across instrumented circuits
  - Better understanding of large flow sources/sinks
  - Performance information for data transfers
- Collaboration between Indiana University, LBNL, and University Hawaii Manoa
- Originally funded by the NSF international program, software is now being deployed domestically as well
- International networks dashboards: http://portal.netsage.global

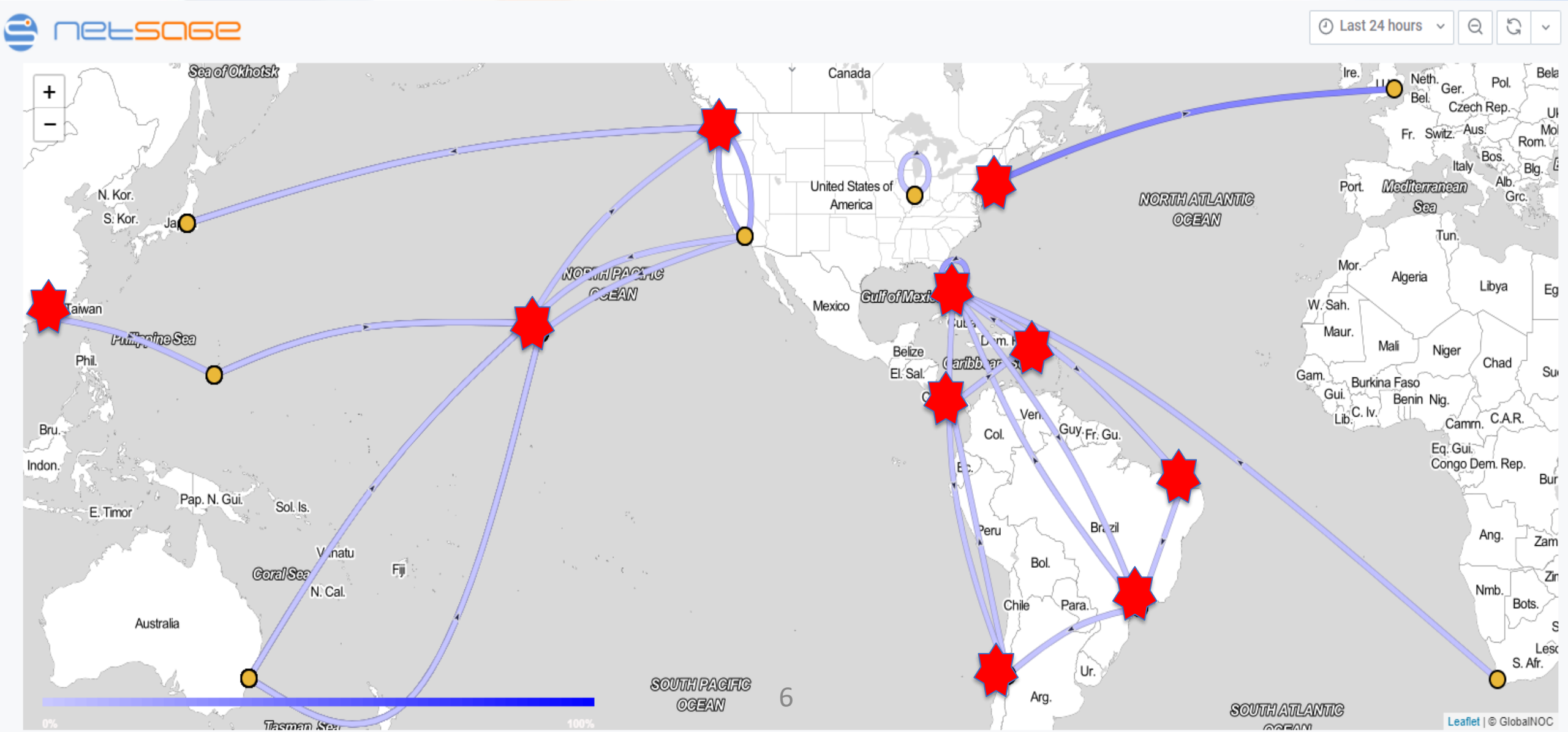
IRNC NetSage Focus on Use Cases

- Bandwidth Dashboard: http://portal.netsage.global
  - How used are the links?
  - Where are congestion points?
- Flow Data Dashboards
  - What are the top sites using the circuits?
  - What are the top sources/destinations for an organization?
  - Who's using my archive?
- Moving towards debugging dashboards
  - What are the flows like between these two orgs?
  - There was a performance spike on my circuit – what was it?
  - Who's transferring a lot of data really slowly?

IRNC NetSage Data Sources

- SNMP data (Passive) - Basic bandwidth data
  - All IRNC Projects
- perfSONAR (Active)
  - IN@IU supported IRNC PS Nesh
- Flow data from routers (Passive)
  - NY (NEAAR), Miami (AmPath/AmLight), Seattle (TP), HongKong (TP), Hawaii-LA (PIREN)
  - PacWave
- Tstat-based traffic analysis for archives (Passive)
  - NCAR, TACC, NERSC, UHawaii Astro

NSF International Circuits – Flow Data Collectors



By Volume

Destination Organization	Total Volume	Largest Flow	# Flows
Boston University	42.2 TB	52.7 GB	42.7 K
New York University	13.7 TB	38.3 GB	986.0
Harvard University	13.7 TB	274.3 GB	9.3 K
Longwood Medical and Academic Area (LMA)	12.3 TB	44.0 GB	7.4 K
Brookhaven National Laboratory	8.5 TB	14.1 GB	7.3 K
University of Hawaii	8.4 TB	28.4 GB	1.5 K
The University of Hong Kong	5.8 TB	60.9 GB	17.3 K
WIDE Project	3.3 TB	21.9 GB	557.0
Yale University	2.6 TB	523.5 GB	4.5 K
Massachusetts Institute of Technology	2.5 TB	53.2 GB	8.2 K

By Rate

Destination Organization	Peak	Average	# Flows
University of Hawaii	8.0 Gbps	1.5 Gbps	1.5 K
Boston University	5.5 Gbps	147.1 Mbps	42.7 K
Brookhaven National Laboratory	3.8 Gbps	222.7 Mbps	7.3 K
Massachusetts Institute of Technology	3.4 Gbps	45.9 Mbps	8.2 K
The University of Hong Kong	3.3 Gbps	21.0 Mbps	17.3 K
Internet2	3.0 Gbps	1.5 Gbps	23.0
Chinese University of Hong Kong (The)	1.5 Gbps	23.4 Mbps	2.8 K
Fermi National Accelerator Laboratory (Fermilab)	1.1 Gbps	120.6 Mbps	297.0
Universidade Federal do Rio Grande do Sul	1.0 Gbps	90.4 Mbps	23.0
New York Genome Center, Inc.	834.8 Mbps	157.4 Mbps	758.0

Top Flows by Organization -- JISC

Individual Flows -- JISC

# Flows	Total Volume	Avg Rate
158,145	134.6 TB	106.6 Mb/s

Flows from Source to Destination

Timestamp	Source Organization	Source Subnet	Destination Organization	Destination Subnet	Total Volume ↓	Rate	Duration
2020-11-02 06:36:34	Jisc Services Limited	163.1.240.x	University of Pennsylvania	128.91.45.x	541.6 GB	50.7 Mb/s	23:45:01
2020-11-02 21:48:29	Jisc Services Limited	193.62.193.x	Yale University	192.31.2.x	523.5 GB	115 Mb/s	10:08:08
2020-11-02 21:57:22	Jisc Services Limited	193.62.193.x	Yale University	192.31.2.x	488.8 GB	142 Mb/s	07:40:20
2020-11-02 21:54:51	Jisc Services Limited	193.62.193.x	Yale University	192.31.2.x	397.2 GB	122 Mb/s	07:13:20
2020-11-02 21:55:54	Jisc Services Limited	193.62.193.x	Yale University	192.31.2.x	360.5 GB	145 Mb/s	05:31:22
2020-11-03 02:10:00	Jisc Services Limited	129.11.190.x	National Aeronautics and Space...	198.120.185.x	295.1 GB	33.9 Mb/s	19:21:09
2020-11-03 00:09:15	Jisc Services Limited	128.232.224.x	Harvard University	140.247.139.x	274.3 GB	195 Mb/s	03:07:10
2020-11-03 00:10:31	Jisc Services Limited	128.232.224.x	Harvard University	140.247.111.x	268.0 GB	191 Mb/s	03:07:22
2020-11-03 00:10:07	Jisc Services Limited	128.232.224.x	Harvard University	140.247.139.x	243.2 GB	173 Mb/s	03:07:10
2020-11-02 02:10:00	Jisc Services Limited	129.11.190.x	National Aeronautics and Space...	198.120.185.x	230.3 GB	21.3 Mb/s	23:59:59
2020-11-07 11:13:49	Jisc Services Limited	193.62.193.x	University of Massachusetts	205.172.168.x	188.4 GB	158 Mb/s	02:39:20
2020-11-03 08:02:11	Jisc Services Limited	128.232.224.x	Harvard University	140.247.139.x	175.9 GB	198 Mb/s	01:58:45
2020-11-02 22:32:08	Jisc Services Limited	193.62.193.x	Yale University	192.31.2.x	169.3 GB	106 Mb/s	03:32:30
2020-11-07 14:22:58	Jisc Services Limited	193.62.197.x	University of Massachusetts	205.172.168.x	149.3 GB	99.9 Mb/s	03:19:16

Detailed Flow Information

Volume	Rate	RTT
523.5 GB	114.8 Mb/s	N/A

Common Information

Timestamp	Sensor	Flow ID	5-Tuple Hash	Protocol	Duration
2020-11-02 21:48:29	NEAAR New York sF...	195EjnUBBtilhIpP9vxG	dc3f9b5e0f32edada3f070e37c8cc5fc1a5c6e5329a2166779288c38201242fe	udp	10:08:08

Source Information

Field	Data
Organization	Jisc Services Limited
Country	United Kingdom
ASN	786
Subnet	193.62.193.x
Port	33001
Science Registry Discipline	BIO.Genomics and Bioinformatics
Science Registry Org Name	European Bioinformatics Institute (EMBL-EBI)
Science Registry Org Abbreviation	EMBL-EBI
Science Registry Project Name	No Data
Science Registry Project Abbreviation	No Data

Destination Information

Field	Data
Organization	Yale University
Country	United States
ASN	29
Subnet	192.31.2.x
Port	34564
Science Registry Discipline	No Data
Science Registry Org Name	No Data
Science Registry Org Abbreviation	No Data
Science Registry Project Name	No Data
Science Registry Project Abbreviation	No Data

Flows by country

- Still a great way to see weird routes
  - One under investigation - traffic from Egypt
    - Flows per country, limit to NEAAR sensor (NY)
  - https://portal.netsage.global/grafana/d/fgrOzz_mk/flow-data-per-country?orgId=2&var-Sensors=NEAAR%20New%20York%20sFlow&var-Country=Egypt
  - Click on a country to see traffic between those 2 countries

Science Registry

- Updated disciplines to match NSF
- https://portal.netsage.global/grafana/d/WNn1qyaiz/flows-by-science-discipline?orgId=2
- Next step will be using WhoIs data to match “Unknown Dest”

Flow Summary Statistics

- See the general state of the IRNC circuits
- https://portal.netsage.global/grafana/d/CJC1FFhmz/other-flow-stats?orgId=2

General Infrastructure Improvements

- Improve dashboard load times by leveraging recent additions to Grafana that allow caching of queries between panels
  - Example: A graph and a table running the same query only need to ask the database once, whereas currently they ask independently
- Migrate perfSONAR data from TSDS to ElasticSearch using push as opposed to pull model
  - Faster availability of data
  - Leverage Elastic query language
  - Alignment with general direction of perfSONAR community

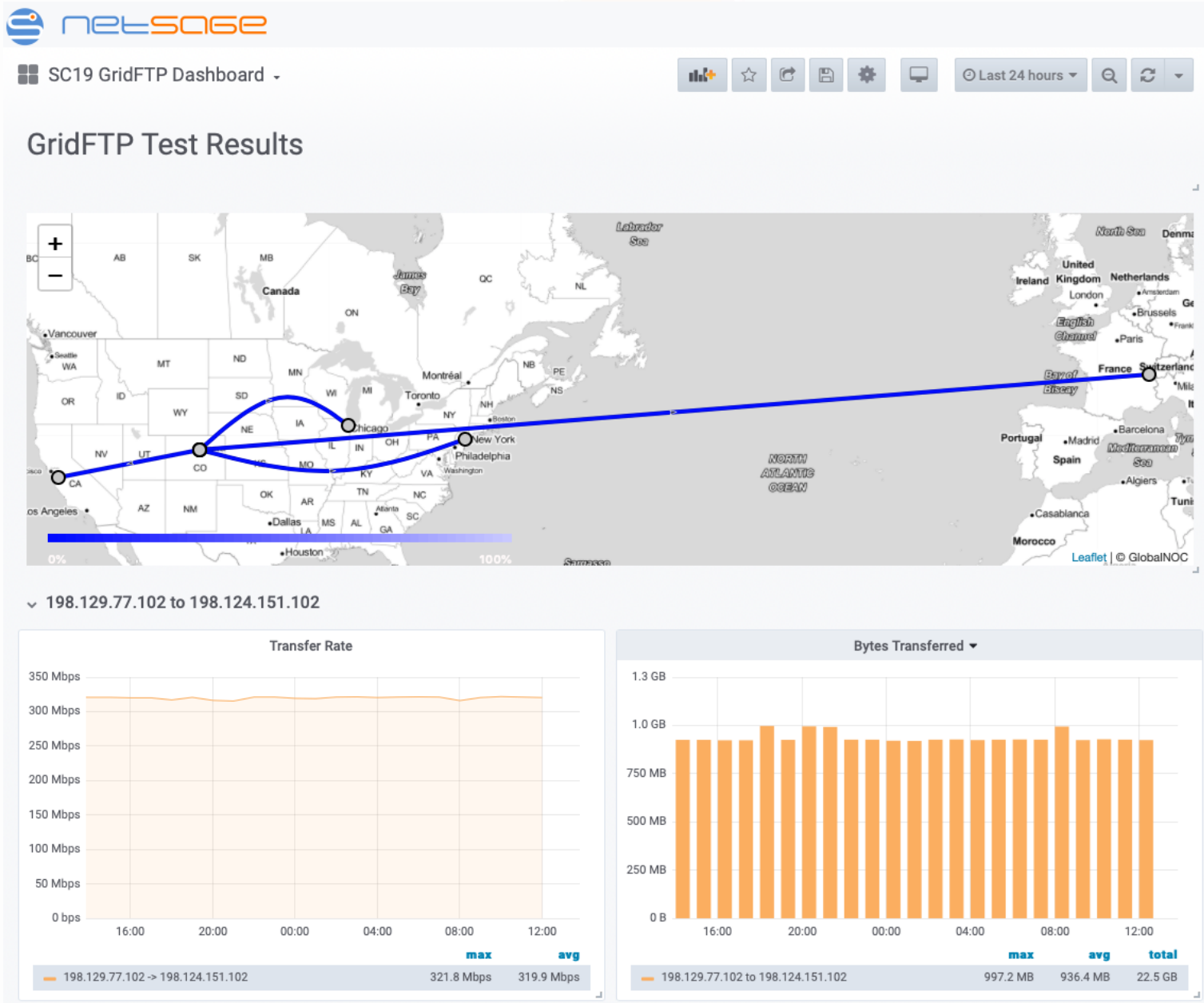
Planned Infrastructure Improvements

Data Roll-ups for large time spans:

- Reduces query times by grouping together flow data once it passes specific age thresholds
- Faster results at the expense of data fidelity
- Not usually an issue when looking for patterns of transfers over long, past time periods
- Recent, individual flow data will still be available

GridFTP and NetSage

- Single point-in-time tests can be useful, but having access to historical data gives useful insights
  - Usage patterns/times of low utilization
  - Locate failing equipment (decreasing transfer performance over time)
  - Real transfer performance with peers vs theorized
- NetSage framework is flexible enough to digest this historical data and display it in an easily accessible visual format
  - Quick visualization vs time consuming log file analysis



NetSage SC19 Visualization

- GridFTP tests between SCinet and ESnet DTNs
- Transfer rate, volume, and locations
- Can easily accommodate data from any DTN running perfSONAR

Acknowledgements

- IN@IU is funded by
  - US NSF award #1450904 for TransPAC4
  - US NSF award #1540933 for NetSage
  - US NSF award #1638863 for NEAAR
  - US NSF award #1826994 for EPOC
- The TransPAC PacificWave 100gb/s network fabric is provided by Pacific Northwest GigaPop